Appendix

Inference for Categorical Data with Excel, JMP, Minitab, SPSS, CrunchIt!, R, and TI-83/-84 Calculators

Inference for Two-Way Tables

Most of the software packages (except Excel and TI calculators) want the data for the two-way table entered in the following format:

Sex	Live with parents	count
Female	Yes	923
Female	No	1703
Male	Yes	986
Male	No	1287

EXCEL

1. Enter the data as a two-way table (the layout used in the text).
2. Compute the expected counts as (row total)×(column total)/(table total) into another two-way table in the spreadsheet. (You can use the Sum(xx:yy) function to find the appropriate totals.)
3. Click **Formulas ➔ More Functions ➔ Statistical ➔ CHISQ.TEST**. Specify the range of cells for both the observed and expected counts.

For a video that shows how to use Excel and the formulas with an example, see the Excel Video Technology Manual: Chi-Square Two-Way Test.

jmp

1. Enter the data as shown above.
2. Click **Analyze ➔ Fit Y by X**.
3. Select and enter the two categorical variables as the response variable and factor variable and the counts as Weights. Even if there is a clear explanatory/response relationship in the data, which variable takes on which role here is really unimportant, but the "response" will become the column variable in the two-way table.
4. Click **OK**. JMP creates the two-way table with marginal distributions in addition to the mosaic plot. The test results you are interested in are labeled "Pearson."

For a video that shows how to use JMP here with an example, see the JMP Video Technology Manual: Chi-Square Two-Way Test.

MINITAB

1. Enter the data as shown above.
2. Click **Stat ➜ Tables ➜ Cross-Tabulation and Chi-square**.
3. Use the drop-down to select whether you have raw or summarized data.
4. Select one of the two categorical variables to display as the row variable and the other as the column variable. Select the "Counts" variable to be the "Frequencies" variable. You can check boxes for row, column, and/or total (joint) distribution percents to be added to the table.
5. Click **Chi-Square**. Click to check the box labeled "Chi-square test." Check other boxes to display other statistics as desired.
6. Click **OK** to return to the main dialog.
7. Click **OK** for the results.

For more information and an example, see the Minitab Video Technology Manual: Chi-Square Two-Way Test.

SPSS

1. Enter the data as shown on page TA11-1. (SPSS really wants individual case data, but that is impractical, as all data you should encounter with this text are summarized.)
2. Click **Data ➜ Weight Cases**. Select to "Weight cases by" the count variable, then click **OK**.
3. Click **Analyze ➜ Descriptive Statistics ➜ Crosstabs**. Select one of the categorical variables to be the row variable and the other for columns.
4. Click **Statistics**. Check the box for the "Chi-square," then **Continue**.
5. If you want SPSS to display the expected counts and/or marginal or joint distributions, click **Cells** and check the box(es) for them.
6. Click **Continue ➜ OK** to perform the test.

For more information and an example, see the SPSS Video Technology Manual: Chi-Square Two-Way Test. The example with summarized data starts about 3:10 into the video (the preceding portion is for raw data).

CRUNCHIT!

1. Enter the data as shown on page TA11-1 (or open a supplied data file).
2. Click **Statistics ➜ Contingency Table ➜ with counts** (unless you have raw data, which is very unlikely).
3. Select one categorical variable for rows, the second to be the column variable, and the third with the observed counts.
4. Click **Calculate**. The output includes both marginal distributions as well as the joint distribution.

For more information (and an example), see the CrunchIt! Help Video: Chi-Square Two-Way Test.

TI-83/-84

1. Press [2nd] [x⁻¹] [▶] [▶] for the Matrix Edit Menu. Select a matrix (usually [A]) in which to enter the observed counts using the arrows. Press [ENTER] to actually enter the matrix editor.
2. Enter the counts across each row, pressing [ENTER] after each entry to move to the next.
3. Press [2nd] [MODE] to exit the editor.
4. Press [STAT] [▶] [▶] for the Tests menu, then select option C:χ^2-Test. If needed, press [2nd] [x⁻¹] to change the matrix names for the observed and expected counts. Press [ENTER] on Calculate.

For more information and an example, see the TI-83/-84 Video Technology Manual: Chi-Square Two-Way Test.

1. With raw data, use the table function to create a table:

   ```
   > Table (datasetname) -> twoway.name
   ```

   where "name" is a name you give the table. Follow that with

   ```
   > chisq.test(twoway.name)
   ```

2. With summarized data, use the command

   ```
   > xtabs (Count ~ rowvar + colvar) -> twoway.name
   ```

   where those are the variables to be treated as rows (or columns) of the output matrix).

   Follow this with

   ```
   > chisq.test(twoway.name)
   ```

3. In either case, to see the expected counts, use the command

   ```
   > results$exp
   ```

For more information and an example, see the R Video Technology Manual: Chi- Square Two-Way Test.

Goodness of Fit

EXCEL

1. Enter the category names in a column (say, A), the observed counts in a column (say, B), and the given probabilities in a column (say, C).
2. Place the cursor in an empty cell (say, B15) and sum the observed counts using the command

   ```
   = sum(B1:Bn)
   ```

   where n is the last row of counts.
3. Compute the expected counts by placing the cursor at the top of an empty column (say, D) using the command

   ```
   =C1*B$15
   ```

4. Drag the cursor to copy the command down until all the expected counts have been calculated.
5. Calculate the parts of the chi-square statistic by placing the cursor at the top of a new column (say, E) using the command

   ```
   =(B1−D1)^2/D1
   ```

6. Drag the cursor to copy the command down until all the chi-square parts have been calculated.
7. Sum the parts for the statistic:

   ```
   =sum(E1:En)
   ```

8. Click **Formulas ➔ More Functions ➔ Statistical ➔ CHISQ.DIST.RT.**
9. Enter the statistic just calculated and the degrees of freedom (categories – 1). Click **OK** for the *P*-value.

JMP

1. Enter the category names in one column and the observed counts in another.
2. Click **Analyze ➔ Distribution**.
3. Select and enter the categorical variable as Y and the counts as Freq.
4. Click **OK**.
5. Click on the red triangle by the variable name and select **Test Probabilities**.
6. Enter the given category probabilities.
7. Click **Done**. Two slightly different chi-square statistics and P-values will be displayed. You are interested in the one labeled "Pearson."

For a video that shows how to use JMP here with an example, see the JMP Video Technology Manual: Chi-Square GOF Test.

MINITAB

1. Enter the category names in one column, the observed counts in another, and the given probabilities in a third.

   Note: If the null hypothesis is that each category has the same probability, you do NOT have to enter the hypothesized probabilities but can select the radio button labeled "Equal proportions" in the main dialog.
2. Click **Stat ➔ Tables ➔ Chi-square Goodness-of-Fit Test (One Variable)**.
3. Click to enter the column with the observed counts and the column with the category names. Select the radio button for "Specific Proportions" and click to enter the column with the given probabilities or select the radio button for "Equal proportions."
4. Click **OK**.

Minitab's default is to provide two charts after the calculated result: the first is a bar graph that compares observed and expected counts; the second shows contributions to the chi-square statistic. If you do not want to see those, click "Graphs" and uncheck the boxes.

For more information and an example, see the Minitab Video Technology Manual: Chi-square GOF Test.

SPSS

1. Enter the category names in one column, the observed counts in another, and the given probabilities in a third.
2. Obtain the sum of the observed counts: Click **Analyze ➔ Descriptive Statistics ➔ Frequencies**. Click to enter the observed counts as the variable. Click **Statistics** and check the box for "Sum." Click **Continue** and **OK**. We'll call the obtained value of the sum Tot (but in what is to come, you will substitute the actual value).
3. Click **Transform ➔ Compute Variable**. Enter a new variable name. Compute the parts of the Chi-square statistic as

$$(\text{Observed} - \text{Tot*Prob})^2/(\text{Tot*Prob}).$$

4. Click **Analyze ➔ Descriptive Statistics ➔ Frequencies**. Click to enter the variable just calculated (the parts of the chi-square) as the variable. Obtain the sum; this is the value of the test statistic.
5. Click **Transform ➔ Compute Variable**. Enter a new variable name and clear the old formula. Enter "1-" in the Numeric Expression box, then select **CDF and Noncentral CDF** in the Function Group box. Double-click **CDF.Chisq** to transfer that shell into the formula box. Enter the value of the test statistic just obtained and the degrees of freedom (categories – 1). Click **OK**. That result is the *P*-value of the test.

CRUNCHIT!

1. Enter the observed counts in a column. You can enter the category names in a second column, but they will not be used.
2. If the null hypothesis is that all categories are equally likely, there is no need for a column of hypothesized probabilities; otherwise, enter those in a third column.
3. Click **Statistics ➔ Chi-squared Goodness-of-Fit**.
4. Click to enter the column of observed counts and the column of hypothesized probabilities (if there is one).
5. Click **Calculate**.

TI-83/-84

1. Use the Statistics List Editor (STAT ENTER) to enter a list of observed counts and a list of hypothesized probabilities (say, L1 = observed and L2 = probabilities).
2. STAT ▶ ENTER for 1-VarStats. Enter the name of the list with the observed counts. The statistic labeled Σx is the total sample size. In the next step, that is referred to as Tot.
3. Calculate a list of expected counts by placing the cursor in the header of an empty list (L3, say). Enter the formula Tot*L2.
4. If you have a TI-84, press STAT ▶ ▶ for the TESTS menu and then arrow down (or up) to option D:x2GOF-Test. Enter the list names for the observed and expected counts and the degrees of freedom (categories – 1). Press ENTER for results.
5. If you have a TI-83, calculate the parts of the chi-square test statistic by placing the cursor in the header of a cleared list (say, L4) and enter the formula (L1-L3)^2/L3 and press ENTER .
6. Obtain the chi-square statistic by summing L4 as above.
7. Obtain the *P*-value: 2nd VARS 8 gives the x2CDF shell. Enter the test statistic value just obtained, infinity, and the degrees of freedom (categories – 1) separated by commas. Practically speaking, any "large" positive number will work as infinity. An example of the command with four categories is shown below.

   x2CDF(2.345,99999,3). Subtract the results from 1 for the *P*-value of the test.

For more information and an example, see the TI Video Technology Manual: Chi-square GOF Test.

R

1. Enter a variable of observed counts and a variable of hypothesized probabilities.
2. Use chisq.test(obs,p=probs) to perform the test. An example is shown below:

```
> act<-c(15,25,40,60)

> pr<-c(0.1,0.2,0.3,0.4)

> chisq.test(act,p=pr)

   Chi-squared test for given probabilities data: act

X-squared = 0.7738, df = 3, p-value = 0.8557
```

For more information and an example, see the R Video Technology Manual: Chi-square Goodness of Fit Test.